

# Comparando métodos de classificação de arritmia em análises de ECG

JULIANA RESPLANDE SANT'ANNA GOMES<sup>1</sup>

<sup>1</sup>UFG – Universidade Federal Goiás  
INF – Instituto de Informática  
CEP 74.690-900 Goiânia (GO)  
julianarsg13@gmail.com

**Resumo:** Este artigo investiga a classificação arritmia cardíaca através de eletrocardiogramas (ECG), comparando duas técnicas: a partir da regressão linear múltipla e a partir do algoritmo de K-MEANS. No caso de estudo feito, a técnica da classificação regressão linear múltipla mostrou-se superior ao algoritmo de K-MEANS, com acurácia superior em 7% em uma amostra do conjunto de dados do MIT-BIH.

**Palavras Chaves:** ECG, classificação, regressão linear, K-MEANS, MIT-BIH.

## 1 Introdução

Um electrocardiograma (ECG) é um exame que capta sinais e mede a atividade elétrica do coração, fornecendo entre outras informações, uma medição da taxa e do ritmo do coração [1, 2]. Quando o ritmo do coração analisado é anormal, caracteriza-se o quadro de arritmia cardíaca, o qual pode ocasionar AVC ou perda cardíaca [3, 4].

O objetivo deste trabalho é comparar duas técnicas principais de classificação de dados: K-MEANS e classificação a partir de regressão linear múltipla, por meio do estudo de caso da análise arritmia cardíaca em ECG.

## 2 Metodologia

Para o presente trabalho, a MIT-BIH Arrhythmia Database foi empregada, na qual contém registros ECG de diversos pacientes. Todos os registros possuem os batimentos normais e extras sistólicos catalogados por uma equipe médica, que serão detalhados na subseção 2.3. Os métodos testados foram: Classificação a partir de Regressão Linear Múltipla e o Algoritmo K-MEANS, os quais serão explanados nas duas próximas subseções 2.1 e 2.2.

### 2.1 Classificação a partir de Regressão Linear Múltipla

A Regressão Linear Múltipla (RLM) é um método estatístico que procura relacionar a variável independente  $X = (X_1 \dots X_n)$  com a variável dependente  $Y$ , por meio de uma função linear  $f(X)$ , para minimizar o erro:

$$Y = a_1X_1 + \dots + a_nX_n \quad (1)$$

Os valores gerados função linear são contínuos. Dadas as classes  $C_1$  e  $C_2$ , para obter a classificação binária  $g(X)$  a partir de uma regressão linear múltipla  $f(X)$  é necessário traçar um limiar  $L$  tal que:

$$f(X) = C_1, \text{ se } f(X) \leq L \quad (2)$$

$C_2$ , caso contrário

### 2.2 Algoritmo K-MEANS

Dado um conjunto de observações  $X = (X_1 \dots X_n)$ , o algoritmo do K-MEANS busca particionar em  $K$  observações com o objetivo de minimizar a variância dentro de cada partição.

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

### 2.3 Cenários experimentais

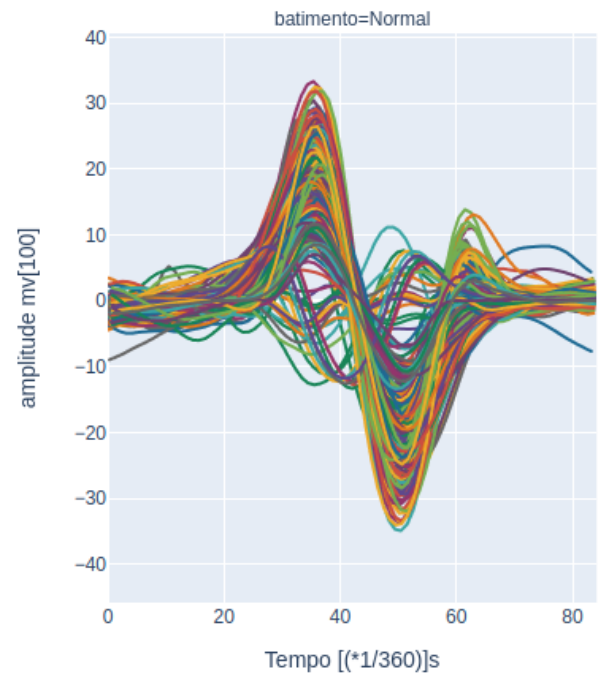
O estudo de caso será focado no paciente anônimo com o identificador “Paciente\_17”. A Tabela 1 a seguir mostra a divisão dos do indivíduo analisado.

**Tabela 1 — Divisão dos dados**

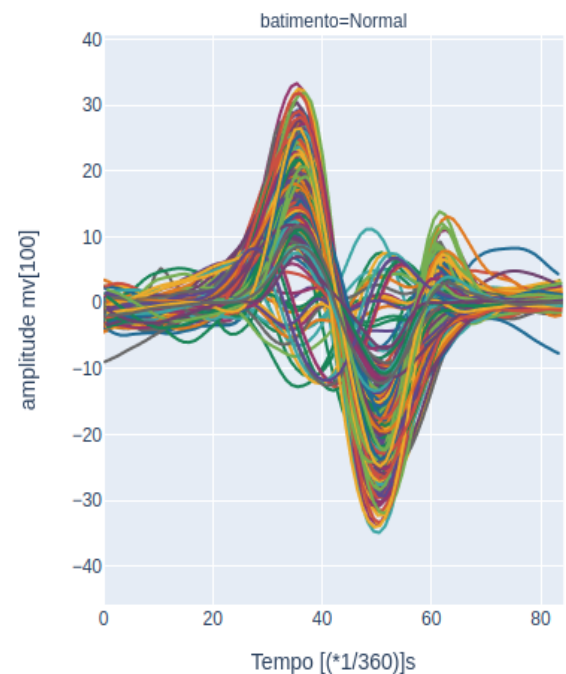
	Batimentos normais	Batimentos extra-sístoles
Treino	294	80
Teste	99	26

Quanto no conjunto de treino e teste, têm-se pelo menos três vezes mais exemplos de batimentos normais do sístoles. Por ser tratar de um caso didático inicial, não houve uma análise envolvendo balanceamento de dados. As Figuras 2 e 3 ilustram os batimentos do conjunto de treino.

No quesito ferramental, foi utilizada a biblioteca Scikit learn<sup>1</sup>, em que foram testadas duas formas de classificação: a partir da regressão linear múltipla e a partir do algoritmo de K-MEANS. Para efeito de reprodutibilidade, ficou-se o limiar de “0” para a Classificação a partir de Regressão Linear Múltipla e a semente de inicialização do algoritmo do K-MEANS como “1”. Será disponibilizado os códigos dos experimentos.<sup>2</sup>



**Figura 2.3.1 - Batimentos normais do conjunto de treino**



**Figura 2.3.2 - Batimentos extra-sístoles do conjunto de treino**

<sup>1</sup> [scikit-learn.org](https://scikit-learn.org)

<sup>2</sup> <https://github.com/jubs12/healthcare-ai>

### 3 Resultados

Os resultados serão divididos em duas subseções. Na primeira subseção, serão ilustradas classificações corretas e incorretas dos classificadores treinados. Posteriormente, será discutido sobre a avaliação no conjunto de teste.

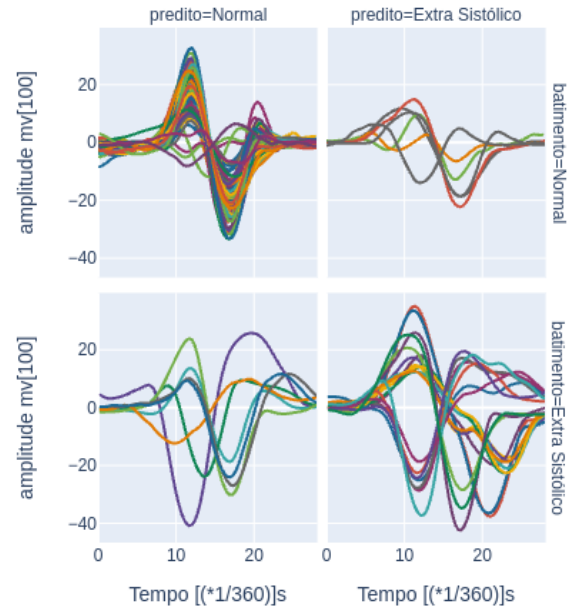
#### 3.1 Demonstração

As figuras 3.1.1 e 3.1.2 ilustram as previsões dos algoritmos de K-MEANS e RLM.

A diagonal principal representa os exemplos em que os batimentos foram classificados corretamente: primeira linha e primeira coluna, batimentos normais; última linha e última coluna, batimentos extra sistólicos.

A diagonal secundária, por sua vez, ilustra os batimentos classificados de forma incorreta: primeira linha e segunda coluna, batimentos extra sistólicos tidos como normais; última linha e primeira coluna batimentos normais preditos como extra-sistólicos.

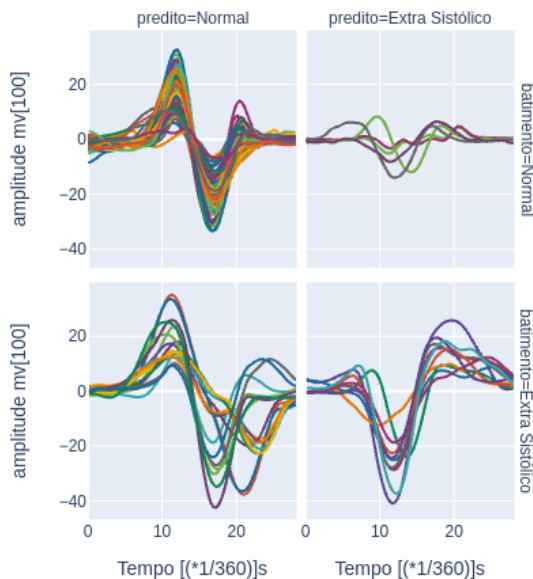
**Figura 3.1.1** - Relação dos batimentos preditos pela K-MEANS no conjunto de teste



**Figura 3.1.2** - Relação dos batimentos preditos pela RLM no conjunto de teste

#### 3.2 Avaliação no conjunto de teste

O algoritmo de K-MEANS obteve 83% de acurácia enquanto o RLM obteve 90% de acurácia. As Tabelas 2 e 3 a seguir ilustram respectivamente os resultados de avaliação por classe em termos de precisão, revocação e F1 para o K-MEANS e a classificação a partir da regressão.



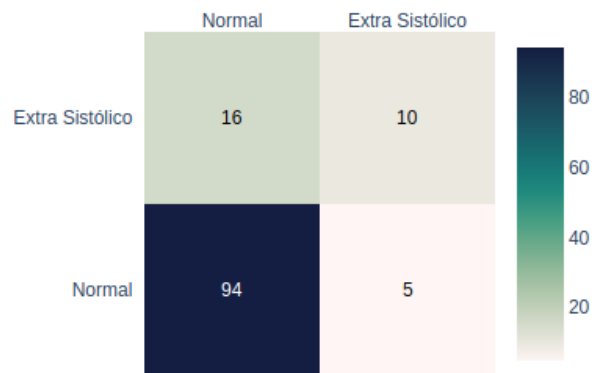
**Tabela 1 — Resultados do Algoritmo de K-MEANS**

	Precisão	Revocação	F1
<b>Normal</b>	0,85	0,95	0,90
<b>Extra Sistólico</b>	0,67	0,38	0,49
<b>Macro</b>	0,76	0,67	0,69
<b>Micro</b>	0,82	0,83	0,81

Em ambas modelos, os resultados dos batimentos Extra-Sistólico são inferiores em relação aos batimentos Normais o que pode ser justificar pelo desbalanceamento dos dados.

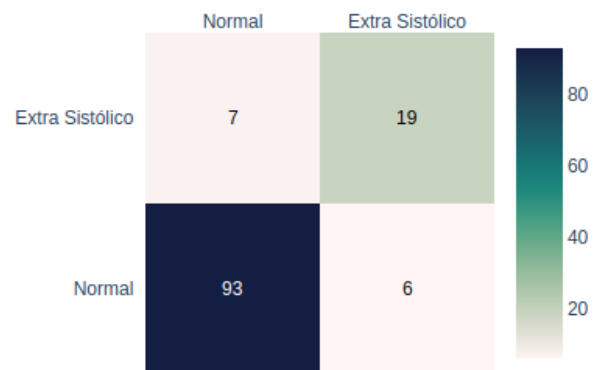
**Tabela 2 — Resultados da Classificação a partir da RLM**

	Precisão	Revocação	F1
<b>Normal</b>	0,76	0,73	0,75
<b>Extra Sistólico</b>	0,93	0,94	0,93
<b>Macro</b>	0,84	0,84	0,84
<b>Micro</b>	0,89	0,90	0,90



**Figura 2.1 — Matriz de confusão do Algoritmo de K-MEANS**

As matrizes de confusão são ilustradas nas Figuras 2.1 e 2.2. A primeira célula das matrizes representa o cenário em que deseja mais evitar: batimentos Extra-Sistólicos classificados como Normais. No algoritmo de K-MEANS, 16 batimentos são classificados desta forma enquanto na Matriz da Classificação a partir da RLM somente sete batimentos.



**Figura 2.2 — Matriz da Classificação a partir da RLM**

#### 4 Conclusões

Obter a classificação a partir da RLM, mostrou-se superior em relação o Algoritmo de K-MEANS com 7% na acurácia. No entanto, é importante notar que foi realizada uma experimentação inicial, sendo relevante a continuação dos experimentos utilizando de técnicas de balanceamento de dados.

## 5 Referências

- [1] Alarsan, F.I., Younes, M. "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms". J Big Data 6, 81 (2019).
- [2] Sampson, Michael, and Anthony McGrath. "Understanding the ECG. Part 1: Anatomy and physiology." British Journal of Cardiac Nursing 10, no. 11 (2015): 548-554.
- [3] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," in IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, May-June 2001, doi: 10.1109/51.932724.
- [4] "Heart Arrhythmia." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 9 Aug. 2020, [www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668](http://www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668).